

A1

1/5/1 (Item 1 from file: 351)  
DIALOG(R)File 351:Derwent WPI  
(c) 2001 Derwent Info Ltd. All rts. reserv.

012877182 \*\*Image available\*\*  
WPI Acc No: 2000-049015/200004  
XRPX Acc No: N00-038385

**Access management device for duplex memory subsystem - assigns logical division of external memory group, according to memory areas in cache provided in controller**

Patent Assignee: HITACHI LTD (HITA )  
Number of Countries: 001 Number of Patents: 001  
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
JP 11312058	A	19991109	JP 98118127	A	19980428	200004 B

Priority Applications (No Type Date): JP 98118127 A 19980428

Patent Details:

Patent No	Kind	Lan	Pg	Min IPC	Filing Notes
JP 11312058	A		8	G06F-003/06	

Abstract (Basic): JP 11312058 A

NOVELTY - The memory in each group is divided into several areas using logical device. Before storing data in each divided memory, the data is stored in cache memory (1300,2300) temporarily in controllers (1000,2000). The controller assigns logical division of memory corresponding to the memory area in each cache.

USE - For duplex memory subsystem in computer network.

ADVANTAGE - High performance is achieved as shift process during failure of memory is avoided. DESCRIPTION OF DRAWING(S) - The figure shows the block diagram of the memory subsystem. (1000,2000) Controllers; (1300,2300) Cache memory.

Dwg.1/5

Title Terms: ACCESS; MANAGEMENT; DEVICE; DUPLEX; MEMORY; SUBSYSTEM ASSIGN;  
LOGIC; DIVIDE; EXTERNAL; MEMORY; GROUP; ACCORD; MEMORY; AREA; CACHE;  
CONTROL

Derwent Class: T01; T03; U21

International Patent Class (Min): G06F-003/06

International Patent Class (Additional): G06F-011/20; G06F-012/08;  
G06F-012/16

File Segment: EPI

1/5/2 (Item 1 from file: 347)  
DIALOG(R)File 347:JAPIO  
(c) 2001 JPO & JAPIO. All rts. reserv.

06370440 \*\*Image available\*\*  
STORAGE SUB-SYSTEM

PUB. NO.: 11-312058 A]  
PUBLISHED: November 09, 1999 (19991109)  
INVENTOR(s): MATSUOTO YOSHIKO  
WRAOKA KENJI  
TAKAOTO KENICHI  
KOBAYASHI MASAAKI  
APPLICANT(s): HITACHI LTD  
APPL. NO.: 10-118127 [JP 98118127]  
FILED: April 28, 1998 (19980428)  
INTL CLASS: G06F-003/06; G06F-011/20; G06F-012/08; G06F-012/16

#### ABSTRACT

PROBLEM TO BE SOLVED: To provide high performance by making it unnecessary to operate any exclusive control or the like under plural controller

control in a normal time even in any connection configuration.

SOLUTION: When a disk array device 5000 in a dual controller constitution and plural port constitution is connected with plural hosts as a redundant constitution, the assignment of a logical volume is divided between controllers 1000 and 2000, and the logical volume is shared by each controller. Even when any host, bus, controller, and cache failure are caused, plural switching means on the occurrence of the failure are provided by a host connection environment, and the switching is not operated as necessary. Then even in the redundant constitution, low overhead in the same level as that in a single constitution can be realized a normal time, and any performance deterioration in the switching processing on the occurrence of each kind of failure can be prevented.

COPYRIGHT: (C)1999, JPO

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-312058

(43) 公開日 平成11年(1999)11月9日

(51) Int.Cl. <sup>8</sup>	識別記号	F I
G 0 6 F 3/06	3 0 4	G 0 6 F 3/06
11/20	3 1 0	11/20
12/08		12/08
		G
		J
12/16	3 1 0	12/16
		3 1 0 J
審査請求 未請求 請求項の数13 O L (全 8 頁)		

(21) 出願番号 特願平10-118127

(22) 出願日 平成10年(1998)4月28日

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 松本 佳子

神奈川県小田原市国府津2880番地 株式会社日立製作所ストレージシステム事業部内

(72) 発明者 村岡 健司

神奈川県小田原市国府津2880番地 株式会社日立製作所ストレージシステム事業部内

(72) 発明者 高本 賢一

神奈川県小田原市国府津2880番地 株式会社日立製作所ストレージシステム事業部内

(74) 代理人 弁理士 小川 勝男

最終頁に続く

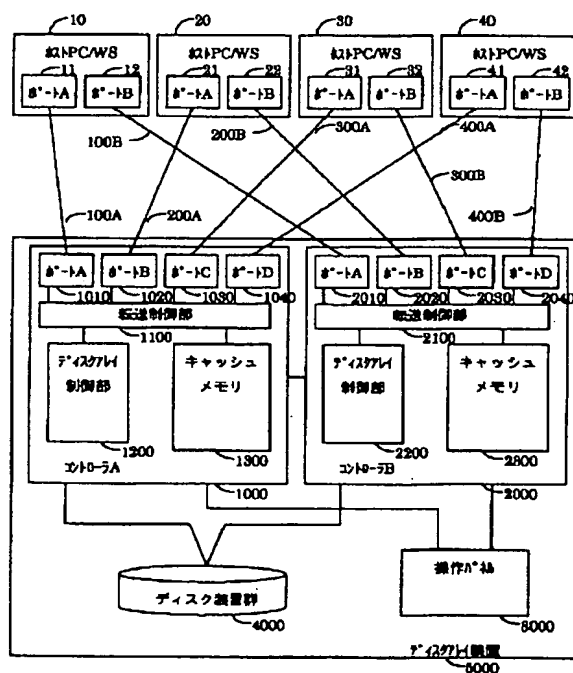
(54) 【発明の名称】 記憶サブシステム

(57) 【要約】

【課題】複数ホスト複数バス構成での冗長構成の制御装置との接続は、論理ボリュームを共有するしないによらず、各部分の障害時、処理の続行は可能であるが、正常時のオーバーヘッドの増加により、シングル構成時との性能差が問題となっている。

【解決手段】デュアルコントローラ構成、複数ポート構成のディスクアレイ装置で複数ホストに冗長構成として接続された時、論理ボリュームの割当てをコントローラ間で分割してコントローラ毎に論理ボリュームを分担する。ホスト、バス、コントローラ、キャッシュ障害時でも、ホスト接続環境により障害時の切替え手段を複数の提供し、場合によっては切替えないようにする。そして、冗長構成においても、正常時もシングル構成と変わらない低オーバーヘッドと、各種障害時の切替え処理における性能劣化を防ぐことを実現する。

図1



## 【特許請求の範囲】

【請求項 1】記憶領域が複数の論理ボリュームに分割される記憶装置群と、

ホストコンピュータに接続され、前記ホストコンピュータと記憶装置との間を転送されるデータを一時的に保存するキャッシュメモリを有するコントローラを複数有し、前記複数のコントローラが前記記憶装置群を制御する記憶サブシステムにおいて、

前記コントローラは、処理を担当する前記論理ボリュームを割り当てられ、前記キャッシュメモリに複数の前記コントローラのための記憶領域を有することを特徴とする記憶サブシステム。

【請求項 2】請求項 1 記載の記憶サブシステムにおいて、

前記コントローラは、複数の前記コントローラのための前記キャッシュメモリの記憶領域を、該コントローラが処理を担当する論理ボリュームのための領域に分割することを特徴とする記憶サブシステム。

【請求項 3】請求項 1 記載の記憶サブシステムにおいて、

前記コントローラが前記ホストコンピュータから処理担当外の論理ボリュームに対する処理要求を受領した時は、該コントローラは、該論理ボリュームの処理を担当していたコントローラの全ての論理ボリュームの処理を担当すること特徴とする記憶サブシステム。

【請求項 4】請求項 1 あるいは 2 記載の記憶サブシステムにおいて、

前記コントローラが前記ホストコンピュータから処理担当外の論理ボリュームに対する処理要求を受領した時は、該コントローラは、該論理ボリュームの処理を担当すること特徴とする記憶サブシステム。

【請求項 5】請求項 1 記載の記憶サブシステムにおいて、

コントローラは、ホストコンピュータから処理担当外の論理ボリュームに対する処理要求を受領した時は、該論理ボリュームの処理担当コントローラに処理要求内容を通信し、通信を受領したコントローラが該論理ボリュームに対する処理を行い、処理結果を要求元コントローラに通信することを特徴とする記憶サブシステム。

【請求項 6】請求項 1 記載の記憶サブシステムにおいて、

コントローラがホストコンピュータから処理担当外の論理ボリュームに対する処理要求を受領した時の処理方法を選択するための入力手段を有することを特徴とする記憶サブシステム。

【請求項 7】請求項 2 記載の記憶サブシステムにおいて、

コントローラ当たりのキャッシュメモリ領域の容量及び、前記論理ボリュームのための領域の容量を指定するための入力手段を有することを特徴とした制御装置を備

えた記憶サブシステム。

【請求項 8】請求項 1 記載の記憶サブシステムにおいて、

複数の前記コントローラのキャッシュメモリのそれぞれに前記複数の論理ボリュームのデータを格納することを特徴とする記憶サブシステム。

【請求項 9】請求項 1 記載の記憶サブシステムにおいて、

10 複数の前記コントローラのキャッシュメモリのそれぞれに前記複数の論理ボリュームのデータを格納するか、前記コントローラのそれぞれが処理を担当する論理ボリュームのデータのみを格納するかを選択する入力手段を有することを特徴とする記憶サブシステム。

【請求項 10】請求項 1 記載の記憶サブシステムにおいて、

前記コントローラに障害が発生したときには、該障害コントローラで担当していた論理ボリュームの処理を正常に動作しているコントローラの担当に切り替えることを特徴とする記憶サブシステム。

20 【請求項 11】請求項 3 記載の記憶サブシステムにおいて、

コントローラがホストコンピュータから処理を担当していない論理ボリュームに対する処理要求を受領したときは、キャッシュメモリ上の管理情報のみの変更で担当の切替えを行うことを特徴とする制御装置を備えた記憶サブシステム。

【請求項 12】請求項 4 記載の記憶サブシステムにおいて、

30 コントローラがホストコンピュータから処理を担当していない論理ボリュームに対する処理要求を受領したときは、キャッシュメモリ上の切替え対象論理ボリュームのうち前記記憶装置群の記憶領域に格納されていないデータを切替え先のコントローラ内のキャッシュメモリ上にコピーして該論理ボリュームの処理を担当するコントローラを切替えることを特徴とする記憶サブシステム。

【請求項 13】請求項 2 記載の記憶サブシステムにおいて、

40 前記コントローラは、前記キャッシュメモリの論理ボリュームのための記憶領域を、該論理ボリュームの負荷に応じて変更することを特徴とした記憶サブシステム。

## 【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、複数台のホスト又は複数のポートを持つ単一のホストと接続する制御装置を備えた記憶サブシステムにおいて、特に詳細には、記憶サブシステムとしての障害時の信頼性及び可用性の向上技術及び障害時の交代系への処理の切替え性能の向上方式技術に関する。

【0002】

50 【従来の技術】コントローラ及びディスク等の記憶装置

に冗長性を持たせたものとして、従来、特開平 4 - 2 1 5 1 4 2 に記載された記憶サブシステムがある。この記憶サブシステムは、2 重の系で構成され、一方の系が現用系として稼働し、他方の系が待機系として稼働している。そして、現用系のディスク装置の記憶情報を両系からアクセス可能な共用ディスク装置を介して予備系のディスク装置に複写すること、あるいは、現用系コントローラ障害時は、予備系のコントローラによって、現用系のディスク装置の記憶情報を抽出可能とすることで、コントローラ及びディスク装置障害時のデータ保全性の向上を計っていた。

【0003】又、従来複数のコントローラで、且つ複数ポート構成で記憶媒体を共有する制御装置は基幹系オープン市場で見られるように、任意の論理ボリュームは複数のバスからアクセスが可能であり、又、全論理ボリュームがどのコントローラからも均一にアクセス可能な構成であった。

【0004】

【発明が解決しようとする課題】コンピュータシステムの大規模化、データ処理の高速化、データの大容量化に伴い、記憶制御装置に対する高性能、高信頼性、可用性の向上が強く望まれている。

【0005】従来技術では、制御装置側の 2 重系、また、キャッシュの 2 重化等により、制御装置内に閉じた信頼性、可用性の向上のみが行われていた。コンピュータシステム全体を考えると、制御装置の 2 重化とともに、ホスト、および制御装置に対するアクセスバスの 2 重化が必須であり、これを実現するためには複数コントローラ構成、複数ポート構成が必須となる。

【0006】しかし、このような構成において従来の基幹系オープン市場のような、全論理ボリュームがどのコントローラからも均一にアクセス可能な構成を実現するためには、複数のコントローラ間でホストのデータを一時格納するキャッシュの管理情報を共有することが必須である。キャッシュの管理情報を共有するには、排他制御を行う為に専用のロック制御、排他情報の管理等をプロセッサで行う必要が有るため、1 つのコントローラ構成での性能に比べ、複数コントローラ構成での性能劣化の割合が大きい。よって信頼性、可用性を向上させるためにその代替として性能を犠牲にするしかなかった。

【0007】システムとして冗長構成をとりながら、冗長構成でないシングル構成時に比べ、性能を維持するためには、どのコントローラからも、全ての論理ボリュームのアクセスが可能でありながら、コントローラ間でのロック制御、排他制御等が不必要であることが必要である。よって、通常のアクセス形態としては、コントローラ単位に論理ボリュームの担当を固定化させる方式が考えられるが、単に固定的な制御では、ホスト、ホスト制御装置間バス、制御装置の障害時、当該担当論理ボリュームへの処理の代替が不可能となり、冗長構成であ

るとは言えない。よって、通常時（正常時）は、各々担当業務を低オーバーヘッドで実現し、障害時は、肩代わりができることが必要である。

【0008】しかし、各コントローラへの接続形態として、1 つのポートで、例えば、デジチェーン接続により複数のホストより接続されている。または、コントローラが複数のポート構成で、各々のポートから異なるホストにて接続されている。さらに、各ホストが、他コントローラ側にも交代バスとして接続しているような構成の時、1 つのホスト側の障害、又はバス障害により、他系コントローラに論理ボリュームを切り替えようとした時、コントローラの担当の切替えとして、他系コントローラの担当論理ボリューム全てを、一括して切り替えてしまう方法では、それ以外のホストが通常処理のままのアクセスを行うことができなくなる。よって、複数のホストと接続する場合には、各々のホスト、バス障害にあった切替え手段が必要となる。

【0009】又、近年フォールトトレラントの強いニーズにより、ホスト側にも、フォールトトレラント機能であるバス切替え機能がサポートされつつある。代表例としては、HP の Alternate Link や、Safe Path（バス切替えソフト）等があげられる。この本機能では、ホスト側が交代バス接続しており、任意バスでの障害を認識すると、自動的に他系のバスへ処理を移し（論理ボリュームを移す）処理を再開する。障害の認識は、タイムアウト等によるものが多く、他系に処理を移した後の最初の I/O で一定時間以上処理が遅延すると、さらに障害と認識される危険性が高い。よって、コントローラの担当の切替え処理には、I/O を受領して認識するわけだが、その後の切替え処理時間は、障害認識されない短い時間で行うことが必須条件となる。

【0010】

【課題を解決するための手段】前述の課題を解決する為に、各々のコントローラが、複数のバス接続ポートとデータを一時的に格納するキャッシュ機構を備え、ホストからのデータを記憶する記憶媒体を共有する複数のコントローラで構成された制御装置で、論理ボリューム（LUN）の担当の指定、及び非担当側コントローラへの I/O 受領時の論理ボリュームの担当の切替え方法の指定、及び論理ボリュームに対するキャッシュの割当て方法の指定、及びキャッシュへの 1 重書き / 2 重書きの指定をする手段と、担当以外の論理ボリュームへの要求時、切り替える制御、及び他系コントローラへ通信する制御を司るプロセッサと、キャッシュへ 2 重書き / 1 重書きの切替えをデータ転送を制御する手段と、キャッシュメモリをコントローラ、さらには、LUN 単位に分割して使用する手段と、LUN 毎の負荷情報、及び LUN の担当により、最適な切替え方法を判断するプロセッサより実現される。

【0011】

【発明の実施の形態】以下、本発明の1実施例を図面を用いて説明する。

【0012】図1は、本実施例に関するディスクアレイ装置を含むシステム全体の1構成例である。図1において、10、20、30、40はホストコンピュータ、5000はディスクアレイ装置であり、各々のホストより、SCSIバス等で接続されている。1000、2000はデュアル構成をとるコントローラ部、4000はホストからのデータを格納するディスク装置群、3000はLUNの指定等を行う操作パネルである。ディスク装置群4000は、アレイ構成であることが多い。

【0013】各コントローラは、バスプロトコル制御を行う4つのポート1010、1020、1030、1040、2010、2020、2030、2040を備え、ホスト側の各ポートに接続される。又、転送制御部1100、2100は、キャッシュメモリ1300、2300とホスト間のデータ転送制御を司り、又1重書き／2重書きの制御も行う。又、ディスクアレイ制御部1200、2200はコントローラA、コントローラB全体を制御している。

【0014】図2にキャッシュメモリの構成を示す。キャッシュメモリ1300、2300は、管理情報1400、2400、コントローラA用エリア1500、2500、コントローラB用エリア1600、2600からなる。各々1400と2400、1500と2500、1600と2600は2重書きされる。キャッシュメモリ1300、2300の管理情報1400、2400は、2重書きにより同一構成であるため、コントローラA側のキャッシュメモリ1400を例にとり管理情報の構成を説明する。管理情報1400は、コントローラ間の通信情報1410と、コントローラ毎のLUN情報1420、キャッシュ内データを管理するデータ管理情報1430からなる。データ管理情報1430は、コントローラ毎に2分割されている。

【0015】キャッシュメモリ1300、2300は信頼性の為、不揮発性であることが望ましい。ホストが、既に冗長構成をとっていて、例えば、ディスクアレイ装置5000を2つ用意し、1方を現行機、他方を予備機として使用しているような場合、敢えてディスクアレイ側での冗長度を要求する必要がある場合、操作パネル3000によるユーザからの指示により、キャッシュを1重書きとして使用する場合は、コントローラA用として1500と1600を、コントローラB用として2500、2600を使用することもできる。この場合、2重書き時と比べ、キャッシュメモリの使用効率率は、2倍となる。この1重書き、2重書きの制御はディスクアレイ制御部1200、2200が、転送制御部1100、2100の転送モードを設定することにより実現される。本キャッシュメモリ1300、2300は、RDキャッ

シュ、及びWRキャッシュとして用いられ高いパフォーマンスを提供することが可能である。WRキャッシュ機能を使用するには、キャッシュメモリ障害、及びコントローラ障害時のデータ保証に為、前記記述のように2重書きすることが必須となっている。

【0016】ホストは、各々2つのポートを持ち、各々のコントローラへ接続される。図1では、ホスト10のポートA：11がSCSIバス100Aを介して、コントローラA：1000内ポートA：100Aへ、ポートB：12-100B-ポートA：2010、以下同様に、別コントローラに接続されている。本接続は、ホスト側の障害、バス障害時、他系に切り替えられる冗長度を持つ構成の代表例である。

【0017】しかし、本ディスクアレイ装置5000の複数ポートに、全て別々のホストに接続されてもよい。この場合ホスト側の冗長度はない。又、ホストが、図1のような2バス構成である時、交代系のバスを同一コントローラ内の別ポートに接続してもよい。本構成をとる場合、コントローラ側の障害時、交代バスがなくアクセスが不能となるが、ホスト側、バス障害時には、同一コントローラへの切替えが可能であり、後記述する切替え処理、及び通信による依頼処理どちらも必要でないため、切替えによる性能劣化がないのが特徴である。

【0018】以下、図1の構成における論理ボリュームの割当てについて説明する。論理ボリュームの割当ては、コントローラ間で排他的に行われる。例えば、コントローラA1000にはLUN0、1、2、3を、コントローラB2000にはLUN4、5、6、7を担当に設定する。担当論理ボリュームは、操作パネル3000で予め設定してもよいし、ディスクアレイ装置の立ち上げ後、受領したコマンドのLUN番号を担当としてもよい。このように、担当LUNを分割し、且つコントローラ単位に使用するキャッシュメモリを分割することにより、マルチコントローラ混成特有の排他制御等が不要となり、高いパフォーマンスを提供することが可能となる。

【0019】コントローラ内の複数のLUNは、複数のホスト間で共有してもよいし、ホスト毎に独立にアクセスしてもよい。この時、コントローラ内のキャッシュメモリの割当てをコントローラ内で一元的に管理するか、各LUN毎に分割して使用するかをユーザが指定することができる。各ホストに独立にLUNを設定した場合、ホスト側のアクセスパターン、ホスト間の能力差等により、任意のホスト配下のLUNのみがキャッシュメモリを占有してしまうと、他ホストからの別LUNに対するパフォーマンスが低下する。本現象を回避するために、予めキャッシュメモリをLUN毎に分割することが可能である。本指定も操作パネル3000により設定することができる。又、LUN毎のアクセスの頻度が時間的に変化するような場合、例えば、日中の業務にはLUN

0, 1を夜間ではLUNの2, 3をメインにアクセスするような場合には、LUN単位の割当てを負荷状態により、ダイナミックに変更することも可能である。LUN単位の割当てを静的に配置するか動的に割り当てるかの指定も操作パネルにて設定することができる。本発明により、マルチホスト接続環境での複数論理ボリュームへのパフォーマンスがホストからのいかなるアクセスパターンにおいてもホストへの均等なサービスを提供することができる。

【0020】次に、LUN毎の割当て方法と、負荷状態にあわせての割当て方法の変更方法を図5を使用し説明する。キャッシュメモリの使用／未使用を管理する最小管理単位をセグメントと称す。データ管理情報はLUN毎に使用可能なセグメント数1431と、現在使用中のセグメント数1432、未使用のセグメントの数1433、各々のセグメント毎のキャッシュの実態ADRや、ホストからのデータADRを管理するセグメント管理情報1434からなる。セグメント管理情報は、セグメントの数と等しい数の情報を持つ。例えば、コントローラA用のキャッシュメモリ1500の大きさが、1,000個のセグメント分だとする。今、コントローラAには、4つのLUNが割り当てられ各々250個セグメントに分割指定されたとする。よって、LUN毎の使用可能セグメント数1431には250が設定される。ホストからI/Oを受領しキャッシュを使用する時、使用中セグメント数1432にくわえ、未使用セグメント数1433から引き、セグメント管理情報1434を設定する。ただし、キャッシュをRDキャッシュをして使用する場合は未使用のまま管理し、WRキャッシュとして使用する場合のみ使用中として管理する。

【0021】また、一定期間で受領したLUN毎のI/O数をカウントしておく。本情報はディスクアレイ制御部に持ってもよいし、キャッシュメモリの管理情報内に持ってもよい。本情報を元にLUN毎の負荷を判定し、各LUN毎の平均I/O数より少ないLUNから多いLUNへ空きセグメントを移行させる。つまり、使用可能セグメント数1431の変更と、未使用セグメント数1433の変更、及びセグメント管理情報の所在を移行することにより実現される。本変更処理は、一定期間毎に行ってもよいし、ホストからのI/Oの負荷の低い時に行ってもよい。本制御により、LUN毎の負荷によりダイナミックにキャッシュの割当てが可能となる。

【0022】次に障害時の切替え処理について説明する。図3にLUNの割当ての1例を示す。本割当て例はコントローラ毎にLUNが分割されているだけでなく、ホスト毎、ポート毎にLUNが独立に設定されている。この構成で、ホスト10のSCSIバス100Aが障害となった時、ホスト側LUN0へのアクセスを交代バス100Bより行う。この時、コントローラBはLUN0はコントローラ毎のLUN情報1420と比較し、非

担当であることを認識する。この時、コントローラA担当分のLUNをすべて一括してコントローラBに切り替えた場合、残りのホスト20, 30, 40がポートA側からLUN1, 2, 3をアクセスする場合の処理もコントローラB側の担当に切り替わることになる。したがって、ホスト20, 30, 40からのLUN1, 2, 3に対するアクセスをコントローラAが受領した場合、非担当LUNに対するアクセスと判断され、もともとこれらのLUNの担当であったコントローラA側にLUNの処理担当を切り替える処理が発生する。このように、コントローラ配下のLUNを一括して切り替える方式では、このようにマルチホスト接続環境では、任意ホスト系障害時、切り替え処理が頻発し、性能が大幅にダウンするといった問題がある。

【0023】本発明では、前記の問題を解決するため、LUN単位の切替えを実現することが可能である。LUN単位の切替えを行うと、上記障害時、コントローラBがLUN0のI/Oを受領した時、切替え対象LUNを当該LUN0のみとすることで、他ホストからのコントローラAへのアクセス時も切り替え処理が発生せず、性能維持が可能である。

【0024】一方、前記の構成で、コントローラAが障害となった場合は、全てのホストがポートBにアクセス権を切り替えてくる為、LUNごとの切り替えより一括切替えの方が、LUNの処理担当を切り替える回数が減少し、LUN処理担当切り替え処理によるの性能劣化を防ぐことができる。

【0025】また、図4にしめすような、LUNの割当て方であった場合、つまり、コントローラ内の各ポート間、つまり、各ホスト間でLUNをシェアして使用する場合、コントローラ側障害時は上記一括切替えにて処理続行が可能であるが、ホスト側障害又は、SCSIバス障害時は上記2つの切替え処理でも、切り替え動作が頻繁に発生してしまい、パフォーマンスが大幅にダウンしてしまう。例えば、ホスト10のSCSIバス100Aが障害となった場合、ホスト10は、コントローラA側での処理を交代バス100Bを使用しコントローラBに発行してくる。つまり、LUN0, 1, 2, 3に対する要求をコントローラBに発行する。

【0026】コントローラBは、非担当LUNのI/Oを受領したことを契機に切り替え処理を行う。一括切替えの場合、コントローラB側に全てのLUNの担当が移行する。その後、他ホスト20, 30, 40からLUN0, 1, 2, 3に対する処理要求がコントローラAを介してあったとき、再び、一括切替えが発生する。すなわち、ホスト10からのI/O及びホスト20, 30, 40からのI/Oの度に一括切替えが起こる。LUN単位切り替えでも、まったく同様な切り替えの繰り返しが起こる。

【0027】前記の問題を解決する為、本発明は非担当

LUNに対するI/Oを受領した時、切り替えない方式も提案している。本方式は、コントローラ間通信情報1410を使用し、相手コントローラに相手コントローラエリアのキャッシュの確保とディスク装置4000との読み出し、書き込みを要求する。相手コントローラは、要求処理が終了すると、コントローラ間通信情報1410にその旨を報告する。そして自コントローラは、対ホストとの転送のみを実行する。本発明により、コントローラ内のLUNを複数ホスト間で共有する構成で、ホスト側やバス障害時の交代バスからの処理の続行を頻繁な切替え動作なしに、実行することが可能である。

【0028】本発明では、一括切り替え処理方式、LUN毎の切替え処理方式、切替えなしに相手コントローラへの依頼処理方式の3つのモードから、ユーザが自らのシステム構成のアクセス環境を考慮して最適な方式を選択することができる。この選択は、操作パネル3000を介して行う。さらに、ユーザがLUN毎の切り替え処理方式を選択した場合でも、コントローラに障害が発生した場合は、一括切り替え方式によりLUN処理担当を切り替えるようにしてもよい。

【0029】又、本発明では、ディスクアレイ装置5000側でI/Oパターンより自動選択する自動モードも設定することができる。自動モード時の選択方法を以下に説明する。

【0030】LUNの割当て時、コントローラ毎にLUNを割り当てた場合、つまり、各ポート毎に設定しなかった場合は、コントローラ内のLUNを各ポート（ホスト）間でシェアするか独立なのかの判断ができない。よって、一定期間の各ポートからのアクセスLUNの種類をカウントし、一定期間後シェアであるか、独立であるかを判断する。独立であれば、LUN単位の切替えを選択するし、シェアであれば、依頼方式を選択する。尚、アクセスパターンが、時間と共に変わる可能性が有る為、さらに一定期間後に上記処理を行う。

【0031】一括切替えは、シングルポートでのデュアル構成で、交代バス構成のシングルホスト構成時等に有効である。又、図1の様な複数ホスト構成で、LUNをシェアしてアクセスしている時にも、ホスト間で切り替えたことを通知する。任意ホストの切替え処理に伴い、他ホストも一括して交代バスに切り替える様な制御が可能なホスト環境においては有効である。

【0032】又、本発明の記憶サブシステムでは、コントローラの各ポート毎に処理を担当するLUNを割り当てるとき、1つのコントローラ内の複数のポートであれば同一のLUNの処理を担当するように割り当てることができる。この場合も、コントローラ内の複数のポートに対するLUNの割り当て方法に基づいて、コントローラがホストコンピュータから処理を担当していないLUNに対する処理要求を受け取ったときに、全LUNの切替え方式か、LUN毎の切替え方式か、依頼方式かを自

動的に選択するようにできる。

【0033】次に、一括切り替え処理の処理手順について説明する。コントローラBが、非担当LUNを受領したとする。この時、コントローラ間通信情報1410にて、その旨をコントローラAに伝える。コントローラAは当該情報により、一括切替えを認識し、現在行っている処理を中断する。中断したら、その旨をコントローラBに、コントローラ間通信情報1410を使用し伝える。コントローラBは、コントローラAからの報告を受けると、コントローラ毎のLUN情報1420の変更を行い、コントローラA配下の全LUNの担当をコントローラBに設定する。又、データ管理情報1430のコントローラA用の管理情報も、全てコントローラB側で管理することになる。この様に、一括切替えの場合は、管理情報の更新のみで切り替えることが可能であり、高速な切替え手段を提供することができる。

【0034】次にLUN毎の切替え方法について説明する。コントローラB側が、非担当LUNを受領し、コントローラAから切り替える場合を説明する。コントローラBが非担当LUNを受領したとする。この時、コントローラ間通信情報1410にて、その旨をコントローラAに伝える。コントローラAは、当該情報により、LUN単位切替えを認識し、現在行っている処理を中断する。中断することにより、コントローラAが、現在キャッシュメモリをアクセスしていないことを保証する。中断したら、その旨をコントローラBに、コントローラ間通信情報1410を使用し伝える。コントローラBは、コントローラAからの報告を受けるとコントローラA用、B用のデータ管理情報1430を一旦コントローラBが管理する。コントローラBは、切り替え対象LUNの使用可能セグメント数1432と、当該セグメントの管理情報1434を参照し、使用中の数だけ、コントローラB側のデータ管理情報内の各LUN毎の未使用セグメント数と、当該セグメント管理情報より、コントローラAの使用可能セグメント内データをコントローラB側の空きセグメントへコピーする。本コピー処理自体は、コントローラB内の転送制御部2100により実現することができる。

【0035】データ自体のコピーと共に、コントローラA側の切替え対象LUNのLUN毎のデータ管理情報を削除し、元々、当該LUNに使用可能セグメント数として割り当てられていたセグメントを他のLUNに分配する。分配されたセグメントは、未使用セグメント数として登録される。又、反対にコントローラB側のデータ管理情報は、切り替え対象LUN情報を追加する。当該LUNに、コピーした数分使用中セグメントとして登録し、使用可能セグメント数は、他LUN内の使用可能セグメント数から一部ずつ移行させ、空きセグメントとして登録する。他LUNの使用可能セグメント数も一部ずつ減少させ、また、コピーする為に割り当てた数の空き



セグメントも空きセグメント数から減少させる。本一連の処理が完了したところで、コントローラBは、コントローラAに処理の再開をコントローラ間通信情報1410を使用し伝える。

【0036】また、本制御を実現するために、キャッシュ内で使用中セグメントの数、つまり、WRキャッシュ機能により、ディスク装置4000に未反映のWRデータの数を管理し、一定量以上に達した場合は、ディスク装置への書き込み処理を最優先に行うことにより、一定量以上に増加しないよう管理する。本制御は、ディスクアレイ制御部1200、2200がLUN毎の使用中セグメント数を一定周期で参照し、一定量以上に達したと判断すると、前記書き込み処理を優先的に処理することにより実現される。本制御により、常に一定量以上の空きセグメントが、キャッシュ内に存在することにより、LUN単位のコピー方式での切り替えが可能となる。本切り替え処理は、キャッシュ間でWRデータのコピー処理が必要となるのでキャッシュ搭載容量、また、その時のWRデータ量により、処理時間が増加する。しかし乍ら、近年、データバスの能力は非常に高くなっており、キャッシュ間のコピー処理も数100MB/Sの能力を持つものが一般的であるため、本切り替え処理時間も数秒程度で完了するため、ホストとの整合性上問題は無い。

【0037】

【発明の効果】本発明によれば、コントローラ当たりホストとのバスプロトコル制御を行う複数のポートを持ち、コントローラ及びキャッシュを2重化した記憶サブシステムにおいて、複数ホスト間との接続において、コントローラ毎に独立に論理ボリュームの割当てを行うことにより、いかなる接続形態においても、通常時、複数コントローラ制御による排他制御等が必要なくなり、シングル構成と比べ、高いパフォーマンスが提供できる。又、ホスト側障害、接続バス障害、コントローラ側ポート障害時、ホストが、交代バスを同一コントローラ内の

別バスに接続している場合には、切り替え処理が必要なく、性能劣化のない交代処理が可能となる。又、別コントローラに交代バスを接続している場合には、コントローラ側障害時にも冗長度があるため、全ての障害において、切り替え処理が可能となる。又、ホストからのLUNへのアクセス形態により、切り替え時間の少ない最適な切り替え手段を提供することができる。

【図面の簡単な説明】

【図1】本発明の実施例である制御装置の構成図である

【図2】本発明の実施例であるコントローラのキャッシュ内のデータの構成図である。

【図3】本発明の実施例である複数ホスト間との論理ボリュームの割当て方の一例であり、ホスト毎、ポートごとに独立した論理ボリュームを割り当てている例である。

【図4】本発明の実施例である複数ホスト間との論理ボリュームの割当て方の一例であり、ホスト間で論理ボリュームを共有している一例である。

【図5】本発明の実施例によるキャッシュメモリ内の管理情報の構成図である。

【符号の説明】

10、20、30、40：ホストコンピュータ

11、21、31、41：ポートA

12、22、32、42：ポートB

100A～400A：SCSIバスA

100B～400B：SCSIバスB

1010～1040、2010～2040：ポートA～D

1100、2100：転送制御部

1200、2200：ディスクアレイ制御部

1300、2300：キャッシュメモリ

1000、2000：コントローラ

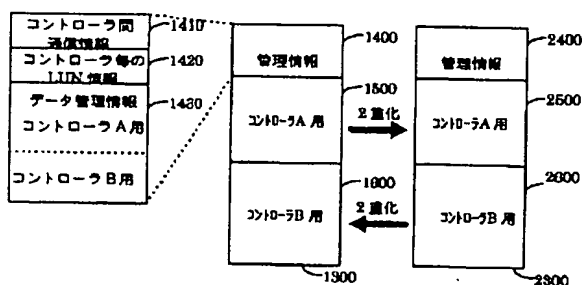
3000：操作パネル

4000：ディスク装置群

5000：ディスクアレイ装置

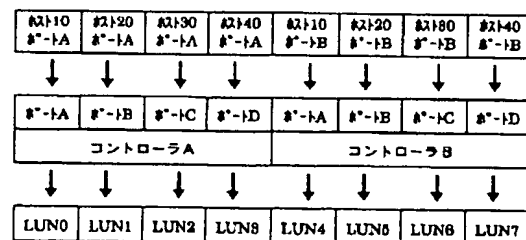
【図2】

図2

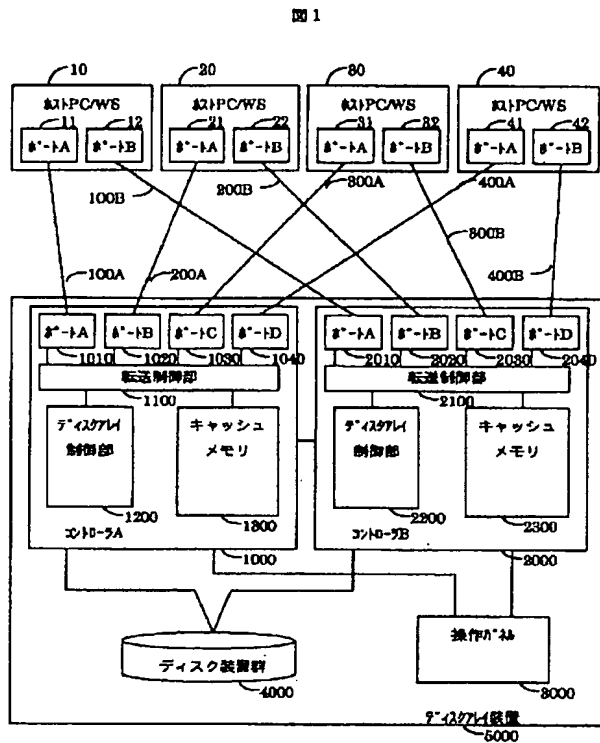


【図3】

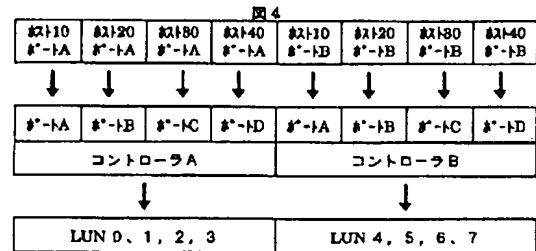
図3



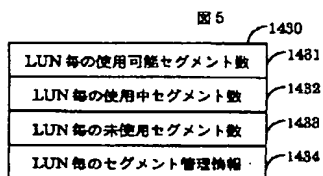
【図 1】



【図 4】



【図 5】



フロントページの続き

(72) 発明者 小林 正明  
神奈川県小田原市国府津2880番地 株式会  
社日立製作所ストレージシステム事業部内